Computer processing of dates outside the twentieth century

by B. G. Ohms

This paper presents practical solutions to problems envisioned in extending computer processing of dates beyond the twentieth century. Many data processing managers are concerned with processing cross-century dates, and in doing so using existing systems, with a minimum of disruption to normal operations. The use of existing date formats can eliminate the need for massive system modifications. Methods of using existing date formats across century boundaries are explained. The use of a format termed the Lilian date format in honor of Luigi Lilio, the inventor of the Gregorian calendar, is introduced. The requirements for an effective date-processing algorithm are presented.

The Gregorian calendar serves us quite well in our day-to-day living. Due to discontinuities in various date divisions, however, it is not readily adaptable to computer programming. This fact becomes more apparent as we approach the new century. Few efficient, easy-to-use functions for manipulating dates have been produced. Also to be considered are the human requirements for ease of use, development, and maintenance. Other considerations include storage costs, efficiency, and adaptability across many different applications and environments.

Some early date-conversion programs were acts of expediency rather than planning, created to solve specific problems rather than for general programming use. Different functions were created at different times. Naming conventions, invocation formats, and implementation methods have often been inconsistent. Programs that provided a day-of-week function were rare. Also rare were programs for deriving a new date by adding or subtracting from another date in a different year. The functions that were available were normally not part of an integrated system for providing compatibility with most of the common date formats. Since documentation was not always created or maintained, individuals were often unaware of what was available. Today, dating format standards are being discussed internationally. The date-processing method presented in this paper is expected to be compatible with any foreseeable international standard date format as well as with the several formats discussed in this paper.

Typically, dates are displayed and stored in the *Julian* and *Gregorian* formats. Concepts and formats of both Julian and Gregorian date formats are discussed later in this paper. Simply stated, the Julian date is the number of the year together with the serial number of the day of that year. The Gregorian date format consists of month, day of month, and year. Many calendars show Julian dates printed in small numerals.

A format termed the *Lilian date format* is presented here as the basis for making date conversions of the

type mentioned earlier. The Lilian format, which may be stored in three bytes of memory, provides the storage capacity for dates well beyond the year 10 000. This format handles processing across century years and other aspects of date conversion not currently adaptable to computer programming. Practical date processing services must provide ease of use for the end user, developer, and maintainer. Such services must also eliminate date ambiguity, achieve excellent performance, and minimize storage.

The two positions traditionally used in both Julian and Gregorian date formats implicitly represent a year within a century. However, this system is inadequate for representing dates in more than one century. For example, it is ambiguous as to whether 03 represents the year 1903 or the year 2003. The Lilian date format avoids the ambiguity by using seven positions for the number of days from the beginning of the Gregorian calendar, October 15, 1582.

Origins of date complexity

Julian calendar. The Julian calendar was established in 46 B.C. by Julius Caesar.[1] It replaced a system of constant adjustments, more for political reasons than for correcting inaccuracies in timekeeping. After a bad start, with the first few leap years being calculated incorrectly, it served quite adequately for many centuries. The Julian calendar is not the same as the Julian date, which was previously defined. The Julian calendar was a time-measuring system, which we now discuss.

The scheme of the Julian calendar was quite simple. The calendar year consisted of 365-day years, with years evenly divisible by four having 366 days. The resulting average year of 365.25 days was slightly longer than the *tropical year*, which is based on the time marked by the passage of equinoxes. That slight difference resulted in a gradual drift of the calendar year. The resulting difficulty in properly setting the date of celebration of Easter caused great concern to the Roman Church. The date for Easter is related to the date for Passover, which is related to the vernal equinox. By the sixteenth century, the discrepancy approximated ten days, and the desire for calendar reform intensified. An artifact of this discrepancy manifests itself today in the differences in dates for Christmas and Easter between the Western Church and the Eastern Church. The latter continues to use the Julian calendar dates as they were at the time of the Gregorian calendar reform. That is, the Eastern Church recognizes the Gregorian calendar, but for certain feast days does not void the ten-day error that existed at the time of the calendar reform.

Gregorian calendar. For the calendar reform, Pope Gregory XIII selected[2] the plan of Aloysius Lilius (1510–1576?),[1] who is also known as Luigi Lilio.[3] This reform, known as the Gregorian calendar, was implemented on Friday, October 15, 1582.[2]

England and her colonies adopted the Gregorian calendar on Thursday, September 14, 1752.

Although use of the Gregorian calendar spread rapidly among Roman Catholic countries, many centuries passed before it was generally used by non-Catholic countries. England and her colonies, including what is now the United States, adopted the calendar on Thursday, September 14, 1752.[3] Many other countries—notably China, Greece, and Russia—did not adopt it until after the beginning of the 20th century. In fact, Russia adopted the Gregorian calendar on two separate occasions, first in 1918, about the same time as many other countries, and again on June 27, 1940.[4] Changes made in Russia in 1929 to avoid the religious associations of the Gregorian calendar were unsuccessful, and it was reimplemented in 1940.

The reform that synchronized the calendar year with the tropical year required the removal of ten days from the calendar. As a result, Friday, October 15, 1582, immediately followed Thursday, October 4, 1582. (An alternate plan would have removed the ten excess days gradually by canceling leap days for 40 years.) As before, every fourth year is a leap year, and, to maintain synchronization, centurial years that are evenly divisible by 400 are also leap years.[2] Thus, for example, 1900 was a common year; the year 2000 will be a leap year, and the year 2100 will start a series of common centurial years again. The result is an average calendar year of 365.2425 days, which closely approximates the tropical year. Despite

this close approximation, by the 44th century,[5] the Gregorian calendar will again differ from the tropical year by one day. Because the tropical year consists of an irrational number of days, no calendar year with an integral number of days can exactly match the tropical year. Not only is there not an integral number of days in a tropical year, but also the length of the day is not constant. That is, the length of the tropical year is also changing gradually.

Discontinuities in the Gregorian calendar scheme cause most of the difficulties with date manipulation.

Algorithms for date processing

Centurian date format. Discontinuities in the Gregorian calendar scheme cause most, if not all, of the difficulties associated with date manipulation. Instances of discontinuities are the following:

- No calendar unit is evenly divisible by weeks.
- The number of days per month varies.
- The number of days per year varies.
- The number of days per century varies.

These facts must be accounted for in calculating the number of days between two dates, calculating a date based on the addition or subtraction of days from another date, and calculating the day of the week for a date. A *centurian date format* (DDDDD, representing the day within a century) works well by giving continuous values within a century. Also, other date formats and day of week are readily calculable from this value. The centurian format is limited to a century and results in discrepancies when a century boundary is crossed. Consideration must be given to century years when making a leap year calculation. The DDDDD format will not span more than 273 years (or 179 years in a two-byte binary format). Also, a standard point of origin for the starting day must be defined.

Lilian date format. The Lilian date format consists of seven positions for the number of days from the beginning of the Gregorian calendar. Day one is Friday, October 15, 1582, and the value is incremented by one for each subsequent day. Based on this format are services supporting 44 experimental functions (or more if the three DMY, MDY, and YMD experimental versions of the Gregorian format are counted) synthesized from eight mnemonics.

Function-naming convention. Functions are named to promote easy recall and to suggest the activity to be performed. In the algorithm and experimental PL/I program discussed in this paper, each function name is synthesized from two 3-letter mnemonic parts. These mnemonic parts are defined as follows:

- CLL: compact Lilian date
- DAY: day of the week
- GRG: Gregorian date
- JUL: Julian date
- LIL: Lilian date
- SGR: short Gregorian date
- SJL: short Julian date
- VAL: validate a date

The first part is a description of the *target*, and the second part is a description of the *source*. VAL (validation) and DAY (day of week) can appear only in the target position of the function.

Arguments presented to any of the conversion functions are always presented in the same order: new date (target); old date (source); range date (control), when required; and Gregorian format [specifier(s)], when required. Table 1 illustrates the format for conversion arguments. In all instances, a return code is set.

The format descriptions in Table 1 refer to the type of data—D for day, M for month, and Y for year, as well as the number of positions (e.g., YY for two year positions)—that the data occupy. The origin of the term "Julian date" for the YYDDD format is given in Reference 6. Nevertheless, dates represented in the Julian format conform to the Gregorian calendar and not to the Julian calendar. The Julian date for Thursday, November 14, 1985, is 85318 (short form) and 1985318 (extended form).

Algorithms. The process of conversion between a Lilian format date and a Julian format date is now described. The algorithms are simpler to calculate by choosing a virtual base date rather than the real one. After the calculations are made, any discrepancies are resolved. In the following algorithms, the num-

Table 1 Examples of conversion function arguments

Argument	Description
target=JULSJL (source, 1925)	Conversion to a Julian (YYYYDDD) date from a short Julian (YYDDD) date within the range 1925–2024
target=JULGRG (source, DMY)	Conversion to a Julian (YYYYDDD) date from a Gregorian (DDMMYYYY) date
target=LILGRG (source, YMD)	Conversion of a Lilian (packed format) date from a Gregorian (YYYYMMDD) date
target=CLLJUL (source)	Conversion of a compact Lilian (binary format) date from a Julian date
target=SGRGRG (source, YMD, MDY)	Conversion of a short Gregorian (YYMMDD) date from a Gregorian (MMDDYYYY) date
CALL VALJUL (source)	Validation of a Julian date
target=DAYSJL (source 1925)	Day of week for a short Julian date

bers in parentheses are results of calculations made on the previously given date, November 14, 1985.

Extended Julian to Lilian. Given an extended Julian date (YYYYDDD), compute the corresponding Lilian date. First, compute the number of days from virtual January 1, 1501, to the start of the year being converted (1985318 Julian).

- Subtract 1501 from the Julian year (484).
- Multiply the difference by 365.25 (i.e., the average number of days per year) (176781.00).
- Truncate the result. The leap day is kept only for full four years (176781).

Then compute the number of Julian calendar (not Julian format) days from October 15, 1582, to the date being converted.

- Subtract 29872, i.e., the number of days between virtual January 1, 1501, and October 15, 1582 (146909).
- Add the Julian days (+318 = 147227).

Next, make the Gregorian calendar adjustments to this value.

- Subtract 1501 from the Julian year (484).
- Divide the difference by 100. One leap year is usually skipped each century (4.84).
- Truncate the result to obtain the number of whole leap days. Partial leap days do not exist (4).
- Subtract the number of whole leap days from the number of Julian calendar days (147223).

Because one out of every four centuries keeps all of its leap years, use the virtual year 1201, which is the beginning of the four-century cycle that contains the year 1582, to compute the number of leap years up to the target date.

- Subtract 1201 from the Julian year. The first four-hundred-year cycle began with virtual year 1201 and ended in the real 1600 (784).
- Divide the difference by 400 because one century leap year per 400 years is kept (1.96).
- Truncate the result because partial leap days do not exist (1).
- Add these leap days back into the adjusted Julian calendar days (147224).

This final result is the number of Gregorian calendar days from the beginning of the Gregorian calendar (October 15, 1582) to the date being converted. This result is the sought-for Lilian date.

Lilian to extended Julian. The purpose of this example is to show the procedure for converting a Lilian date to an extended Julian date. First, create a virtual Lilian date, with a starting point of January 1, 1201. Convert this result into a Julian calendar (not Julian format) date. Then convert the Julian calendar date to a Julian format date. The procedure is as follows (147224 Lilian):

- Add 139444 to the Lilian date. This is the number of days from virtual January 1, 1201 (the start of a 400-year cycle) to October 15, 1582. The result is a pseudo-Lilian date (286668).
- Divide the pseudo-Lilian date by 36524.25, the average number of days per century (7.85).
- Truncate the result to obtain the number of century leap days (7).

- Add the number of century leap days to the pseudo-Lilian date (286675).
- Divide the number of century leap days by 4 (1.75).
- Truncate the result to obtain the number of four-century leap days (1).
- Subtract the number of four-century leap days from the pseudo-Lilian date, because they are already included (286674).

The result is the number of full Julian calendar days from January 1, 1201, to the date being converted. We now convert this to an extended Julian format date.

- Divide full Julian calendar days by 365.25. This usually gives one less than the year for the date being converted (784.87).
- If there is no remainder from the division, subtract one from the quotient; otherwise, truncate the quotient to get the pseudo-Julian prior year (784).
- Multiply the pseudo-Julian prior year by 365.25 to determine the number of annual Julian calendar days prior to the year for the date being converted (286356.00).
- Truncate the number of annual Julian calendar days to eliminate partial leap days (286356).
- Subtract the number of truncated annual Julian calendar days from the full number of Julian calendar days to obtain the number of current Julian calendar days in the year of the date being converted (318).
- Add 1201 to the pseudo-Julian prior year to obtain the real Julian year, i.e., 1200 for years prior to 1200 plus one for the current year (+784 = 1985).
- Multiply the real Julian year by 1000 to put it into its proper Julian format position (1985000).
- Add the current Julian calendar days to the result to complete the Julian format date from the Lilian format date (1985318).

Ease of conversion

Accommodating end users. End users usually enter two digits for the year in a date and understand the ambiguity that this represents. Therefore, even at the turn of the century, to avoid adverse user reaction, programs must continue to function with only two digits for year. The inference of the year 1997 from 97 and 2003 from 03 must continue. For the exceptional case where the correct meaning could be 1897 and 1903, entry of all four digits may be required.

The month-day-year and day-month-year formats are ambiguous. Therefore, it might be advisable to continue presenting the date in its conventional U.S. format with a parenthetical explanation of the format—such as (MM/DD/YY)—to avoid the ambiguity.

It is not necessary to change date formats in files, because it is possible to change the programs only.

However, it may be necessary to provide a conversion function that receives a definition of the implied century as a parameter. An excellent way to do this unambiguously is to specify a year as the desired starting point of a 100-year range. For example, if the starting year for the range is specified as 1925, dates with year digits of 25 through 99 would be between 1925 and 1999, and dates with year digits of 00 through 24 would lie between 2000 and 2024.

Accommodating systems support. The conversion of isolated files to new date formats presents a rather trivial problem. In most cases, however, it is not possible to isolate the process. All programs that access the modified data must be changed simultaneously. In some large systems, literally thousands of programs may be involved. In these large systems, it may be prudent to avoid the cost and risk of massive changes in a short period of time.

It is not necessary to change date formats in files, because it is possible to change the programs only, so that the implied century in a date is recognized. Of course, in the vast majority of cases, that is exactly what does take place. Dates familiarly and implicitly exist within the 100-year range beginning with 1900 or 1901. Thus it is necessary merely to modify the programs so that the 100-year range starts at a later date. A beginning date set eighty years prior to the current systems date may be a reasonable convention. This is well within the range now in use.

The significant feature of this approach is that everything need not be modified at once. The modifications may be made over a period of years during

normal program maintenance. Of course, as systems are maintained or replaced, it would be practical to implement full information date formats. Where systems contain dates that span a range of more than 100 years, the century must already have been carried. In the rare event that this is not true, immediate conversion is unavoidable. Fortunately, in most applications, we can deal with centuries as exceptions rather than as a common problem.

Computational considerations

Storage. The two main considerations pertaining to storage are (1) the cost of storing large quantities of data; and (2) the computational cost of converting records within computer files to larger date-field sizes. When millions of dates are stored, as they are in most business systems, every additional byte required to save a single date multiplies to millions of additional bytes of storage. The programming necessary to accommodate larger date-field sizes in records further complicates date conversion.

Putting bounds on data-storage costs at reasonable levels and reducing conversion complications are both achievable. The allocation of additional space to record four positions for year, rather than the traditional two, is not the only possibility. Many systems currently store dates internally in packed Julian format, requiring three bytes of storage. In packed format, a Lilian date or an extended Julian date (YYYYDDD) both require four bytes of storage. However, if a binary format were to be adopted, either form of date could be stored in the three bytes used at present.

A more restrictive situation exists for systems in which dates are stored in two bytes instead of three. These systems use a binary format to record centurian or other forms of dates. In the near term for these systems, it may be necessary to continue storing dates in only two bytes. This cannot be accommodated with a Julian format. However, it is possible to store a range of dates by taking advantage of the continuous characteristic of a Lilian date.

Using the Lilian date system, any date in a selected range of 179 years may be stored as follows. Assume that the selected range is January 1, 1901, through December 31, 2079. Find the Lilian value of December 31, 1900. Subtract this value from the Lilian value of the date to be stored. Convert the result to a 16-bit (two-byte) binary value and store the result. Reverse the process to restore the two-byte date to Lilian format. Continuing to store dates in two bytes should be considered only where the programming cost of increasing record sizes in an existing system is prohibitive. The decreasing cost of storage makes the use of the full Lilian or Julian format a practical possibility when an application is being modified.

In the experiments on which this paper is based, the three-byte Lilian format for data storage has been found to be preferable. Internally, in computer use, the continuous nature of the Lilian date accommodates the types of processing that normally take place within a program. In addition, for all three storage sizes, a consistent format (i.e., the all-Lilian format or the quasi-Lilian format) is maintained.

The continuous nature of the Lilian date accommodates the types of processing that normally take place.

Flexibility. In our experiments, the IBM System 360/370 assembly language was used for coding the date functions. However, the method is easily adaptable to other programming languages such as COBOL, PL/I, and RPG. The experimental functions were also made re-entrant for flexibility within on-line environments.

Validity. Ideally, the validation of a date is necessary only at the point of its manual entry into a system. Experience teaches us that it is not unusual for an interfacing system to provide invalid dates. Therefore, all dates passed to conversion functions must be validated. When an invalid date is encountered by the experimental system, a null value is returned to the invoking program and a return code is set to indicate the nature of the invalid condition.

Efficiency. Date conversions are used heavily within certain applications. Because of the impact on a single system or an installation, efficiency must be inherent in date-conversion programs. Storage requirements for external display and internal calculations by computer programs are expected to mo-

tivate a high volume of conversions. Widely used programs in high-volume environments must perform well and not consume excessive amounts of storage. A date-conversion program that is good in all other ways will not be widely used if it is an operational bottleneck.

For the program discussed in this paper, written in System 360/370 assembly language and implemented as a set of PL/I functions, the experimental results indicated good efficiency. The longest execution paths, including PL/I prologue, validation, conversion, and PL/I epilogue, are a little more than one hundred machine-language instructions. Although the functions appear to the invoking PL/I program to be separate programs, they are all actually provided within a single program that requires less than 4K bytes of storage.

Experimental results indicated good efficiency.

Concluding remarks

Programmers require date-processing functions that effectively handle applications for both the present and the future. For the present, it is sufficient to validate source dates, to convert from one traditional date format to another, and to perform addition or subtraction operations involving dates. For the future, additional functions are required that support processing restricted only by the limitations of the Gregorian calendar. These functions must be fully compatible with existing date-format and record-size restrictions. Massive conversion efforts should not be required to process and store dates outside the twentieth century. Also, end users should be able to continue to use existing two-digit date formats when interacting with computer systems. In all cases the programs that provide these services must be reliable, efficient in the use of both processor and storage, and flexible in application.

Programs that embody all these qualities have been written and tested experimentally. The application as written requires less than 4K bytes of storage and has an average execution path of fewer than 100 machine language instructions. Re-entrancy and the possibility of multilanguage implementation indicate excellent flexibility. This approach presents a practical method of processing dates that is compatible with any dating format standard.

Acknowledgments

No significant work is done in isolation. Over the years, I have experienced the inspiration, assistance, and patience of many individuals. My colleagues Tom Gauthier and Jim Willard first sparked my interest in date processing several years ago. Recently Alex Chang, Barb Henderson, Jack Henriksen, Fred Lange, Bob Lord, Libby Ross, Karen Seabury, and Billy Shih have patiently served as a sounding board, made helpful suggestions, and have been active supporters. Sandy Mink provided valuable editorial support. Many unnamed others have been involved in my experiment, including every member of my family. Their support is also greatly appreciated.

Cited references

1. G. Moyer, "Luigi Lilio and the Gregorian reform of the calendar," *Sky and Telescope* 64, No. 11, 418–419 (November 1982).
2. J. J. Bond, *Handy Book of Rules and Tables for Verifying Dates Within the Christian Era*. Russell & Russell, a Division of Atheneum House, Inc., New York (1966).
3. *The New Encyclopedia Britannica*. Encyclopedia Britannica, Inc., Chicago (1980), p. 602.
4. F. Parise, Editor, *The Book of Calendars*. Facts on Life, Inc., New York (1982).
5. G. Moyer, "The Gregorian calendar," *Scientific American* 246, No. 5, 144–152 (May 1982).
6. M. A. Covington, "A calendar for the ages," *PC Tech Journal* 3, No. 12, 136–142 (December 1985). Joseph Justice Saliger (1540–1609) named the Julian date in honor of his uncle Julius.

General references

G. Moyer, "Astronomical scrapbook—Notes on the Gregorian calendar reform," *Sky and Telescope* 64, No. 12, 530–533 (December 1982).

International Organization for Standardization, "Writing of calendar dates in all-numeric form," Reference No. ISO 2014-1976(E) (April 1976).

International Organization for Standardization, "Information processing interchange—Representation of ordinal dates," Reference No. ISO 2711-1973(E) (January 1973).

Bruce G. Ohms *IBM Information Systems Group, 301 Merritt 7, Norwalk, Connecticut 06856.* Mr. Ohms joined IBM in 1967 and is currently a senior programmer/analyst working in the area of applications systems development. Prior to his current assignment, he has held a variety of positions in programming, systems design, analysis, and development center implementations. He received an A.A.S. degree in data processing from Belleville Area College, Belleville, Illinois, and he attended Yale University.

Reprint Order No. G321-5274.